



IFW16

## RAW SEQUENCE LISTING

DATE: 09/28/2004

PATENT APPLICATION: US/09/938,901A

TIME: 10:38:43

Input Set : A:\11283-013001.ST25.txt

Output Set: N:\CRF4\09282004\I938901A.raw

3 <110> APPLICANT: Kuramitsu, Seiki  
 4 Yokoyama, Shigeyuki  
 6 <120> TITLE OF INVENTION: Novel DNA Repair Enzymes, Nucleic Acids Encoding DNA Repair  
 7 Enzymes And Methods of Using Them  
 9 <130> FILE REFERENCE: 11283-013001/PH1261 US  
 11 <140> CURRENT APPLICATION NUMBER: US 09/938,901A  
 12 <141> CURRENT FILING DATE: 2001-08-24  
 14 <150> PRIOR APPLICATION NUMBER: JP 47762/2001  
 15 <151> PRIOR FILING DATE: 2001-02-23  
 17 <160> NUMBER OF SEQ ID NOS: 65  
 19 <170> SOFTWARE: PatentIn version 3.3  
 21 <210> SEQ ID NO: 1  
 22 <211> LENGTH: 975  
 23 <212> TYPE: DNA  
 24 <213> ORGANISM: Thermus thermophilus  
 27 <220> FEATURE:  
 28 <221> NAME/KEY: CDS  
 29 <222> LOCATION: (1)..(975)  
 31 <400> SEQUENCE: 1

32 atg gag gcc tgg cgg aaa gcc ctc ctc gcc tgg tac cgg gaa aac gcc	48
33 Met Glu Ala Trp Arg Lys Ala Leu Leu Ala Trp Tyr Arg Glu Asn Ala	
34 1 5 10 15	
36 cgc ccc ctc ccc tgg cgg ggg gag aag gac cct tac cgc gtc ctg gtc	96
37 Arg Pro Leu Pro Trp Arg Gly Glu Lys Asp Pro Tyr Arg Val Leu Val	
38 20 25 30	
40 tcc gag gtc ctt ctg cag cag acc cgg gtg gag cag gcc ctc ccc tat	144
41 Ser Glu Val Leu Leu Gln Gln Thr Arg Val Glu Gln Ala Leu Pro Tyr	
42 35 40 45	
44 tac cgc cgc ttt ctg gag cgc ttt ccc acc ctg aag gcc ctg gcc gcg	192
45 Tyr Arg Arg Phe Leu Glu Arg Phe Pro Thr Leu Lys Ala Leu Ala Ala	
46 50 55 60	
48 gct tcc ctg gaa gag gtc ctt agg gtc tgg cag ggg gcg gcc tac tac	240
49 Ala Ser Leu Glu Glu Val Leu Arg Val Trp Gln Gly Ala Gly Tyr Tyr	
50 65 70 75 80	
52 cgg cgg gcg gaa cac ctc cac cgc ctg gcc cga agc gtg gag gag ctt	288
53 Arg Arg Ala Glu His Leu His Arg Leu Ala Arg Ser Val Glu Glu Leu	
54 85 90 95	
56 ccc ccg agc ttc gcc gag ctt cgg ggg ctt cct ggt ctc ggg cct tac	336
57 Pro Pro Ser Phe Ala Glu Leu Arg Gly Leu Pro Gly Leu Gly Pro Tyr	
58 100 105 110	
60 acc gcg gcg gcg gtg gcc tcc atc gcc ttc ggg gag cgg gtg gcg gcg	384
61 Thr Ala Ala Ala Val Ala Ser Ile Ala Phe Gly Glu Arg Val Ala Ala	
62 115 120 125	

## RAW SEQUENCE LISTING

DATE: 09/28/2004

PATENT APPLICATION: US/09/938,901A

TIME: 10:38:43

Input Set : A:\11283-013001.ST25.txt

Output Set: N:\CRF4\09282004\I938901A.raw

```

64 gtg gac ggg aac gtc cgg agg gtc ctc tcc cgc ctc ttc gcc cgg gaa      432
65 Val Asp Gly Asn Val Arg Arg Val Leu Ser Arg Leu Phe Ala Arg Glu
66      130                      135                      140
68 agc ccc aag gag aag gag ctt ttc gcc ctc gcc cag ggc ctc ctc ccc      480
69 Ser Pro Lys Glu Lys Glu Leu Phe Ala Leu Ala Gln Gly Leu Leu Pro
70 145                      150                      155                      160
72 gag ggc gtg gac ccg ggg gtg tgg aac cag gcc ctc atg gag ctc ggg      528
73 Glu Gly Val Asp Pro Gly Val Trp Asn Gln Ala Leu Met Glu Leu Gly
74                      165                      170                      175
76 gcc acg gtc tgc ctg ccg aaa cgg ccc cgt tgc ggg gcc tgc ccc cta      576
77 Ala Thr Val Cys Leu Pro Lys Arg Pro Arg Cys Gly Ala Cys Pro Leu
78                      180                      185                      190
80 ggg gcc ttc tgc cgg ggg aag gag gcc ccc ggg cgc tac ccc gcg ccc      624
81 Gly Ala Phe Cys Arg Gly Lys Glu Ala Pro Gly Arg Tyr Pro Ala Pro
82                      195                      200                      205
84 agg aag cgc cgg gcg aag gag gag cgc ctc gtc gcc ctc gtc ctc ctc      672
85 Arg Lys Arg Arg Ala Lys Glu Glu Arg Leu Val Ala Leu Val Leu Leu
86      210                      215                      220
88 ggg cgg aag ggg gtg cac ctg gaa agg ctt gag ggg cgc ttc cag ggc      720
89 Gly Arg Lys Gly Val His Leu Glu Arg Leu Glu Gly Arg Phe Gln Gly
90 225                      230                      235                      240
92 ctc tac ggc gtc ccc ctc ttt ccc cct gag gag ctt ccc ggg cgg gag      768
93 Leu Tyr Gly Val Pro Leu Phe Pro Pro Glu Glu Leu Pro Gly Arg Glu
94                      245                      250                      255
96 gcg gcc ttc ggg gtg agg tct agg ccc cta ggc gag gtg cgc cac gcc      816
97 Ala Ala Phe Gly Val Arg Ser Arg Pro Leu Gly Glu Val Arg His Ala
98                      260                      265                      270
100 ctc acc cac cgg agg ctt cgc gtg gag gtg cgg ggg gcc ctt tgg gaa      864
101 Leu Thr His Arg Arg Leu Arg Val Glu Val Arg Gly Ala Leu Trp Glu
102                      275                      280                      285
104 ggg gag ggg gag gac ccc tgg aag agg ccc cta ccc aag ctc atg gag      912
105 Gly Glu Gly Glu Asp Pro Trp Lys Arg Pro Leu Pro Lys Leu Met Glu
106      290                      295                      300
108 aag gtg ctc cgc aag gcg ctt ccc ctc ctc gct cat gcg ggc gta gtc      960
109 Lys Val Leu Arg Lys Ala Leu Pro Leu Leu Ala His Ala Gly Val Val
110 305                      310                      315                      320
112 ccc ctc ccg gac gca
113 Pro Leu Pro Asp Ala
114                      325
117 <210> SEQ ID NO: 2
118 <211> LENGTH: 325
119 <212> TYPE: PRT
120 <213> ORGANISM: Thermus thermophilus
122 <400> SEQUENCE: 2
124 Met Glu Ala Trp Arg Lys Ala Leu Leu Ala Trp Tyr Arg Glu Asn Ala
125 1                      5                      10                      15
128 Arg Pro Leu Pro Trp Arg Gly Glu Lys Asp Pro Tyr Arg Val Leu Val
129      20                      25                      30
132 Ser Glu Val Leu Leu Gln Gln Thr Arg Val Glu Gln Ala Leu Pro Tyr

```

## RAW SEQUENCE LISTING

DATE: 09/28/2004

PATENT APPLICATION: US/09/938,901A

TIME: 10:38:43

Input Set : A:\11283-013001.ST25.txt

Output Set: N:\CRF4\09282004\I938901A.raw

```

133          35          40          45
136 Tyr Arg Arg Phe Leu Glu Arg Phe Pro Thr Leu Lys Ala Leu Ala Ala
137          50          55          60
140 Ala Ser Leu Glu Glu Val Leu Arg Val Trp Gln Gly Ala Gly Tyr Tyr
141 65          70          75          80
144 Arg Arg Ala Glu His Leu His Arg Leu Ala Arg Ser Val Glu Glu Leu
145          85          90          95
148 Pro Pro Ser Phe Ala Glu Leu Arg Gly Leu Pro Gly Leu Gly Pro Tyr
149          100          105          110
152 Thr Ala Ala Ala Val Ala Ser Ile Ala Phe Gly Glu Arg Val Ala Ala
153          115          120          125
156 Val Asp Gly Asn Val Arg Arg Val Leu Ser Arg Leu Phe Ala Arg Glu
157          130          135          140
160 Ser Pro Lys Glu Lys Glu Leu Phe Ala Leu Ala Gln Gly Leu Leu Pro
161 145          150          155          160
164 Glu Gly Val Asp Pro Gly Val Trp Asn Gln Ala Leu Met Glu Leu Gly
165          165          170          175
168 Ala Thr Val Cys Leu Pro Lys Arg Pro Arg Cys Gly Ala Cys Pro Leu
169          180          185          190
172 Gly Ala Phe Cys Arg Gly Lys Glu Ala Pro Gly Arg Tyr Pro Ala Pro
173          195          200          205
176 Arg Lys Arg Arg Ala Lys Glu Glu Arg Leu Val Ala Leu Val Leu Leu
177          210          215          220
180 Gly Arg Lys Gly Val His Leu Glu Arg Leu Glu Gly Arg Phe Gln Gly
181 225          230          235          240
184 Leu Tyr Gly Val Pro Leu Phe Pro Pro Glu Glu Leu Pro Gly Arg Glu
185          245          250          255
188 Ala Ala Phe Gly Val Arg Ser Arg Pro Leu Gly Glu Val Arg His Ala
189          260          265          270
192 Leu Thr His Arg Arg Leu Arg Val Glu Val Arg Gly Ala Leu Trp Glu
193          275          280          285
196 Gly Glu Gly Glu Asp Pro Trp Lys Arg Pro Leu Pro Lys Leu Met Glu
197          290          295          300
200 Lys Val Leu Arg Lys Ala Leu Pro Leu Leu Ala His Ala Gly Val Val
201 305          310          315          320
204 Pro Leu Pro Asp Ala
205          325
208 <210> SEQ ID NO: 3
209 <211> LENGTH: 1998
210 <212> TYPE: DNA
211 <213> ORGANISM: Thermus thermophilus
214 <220> FEATURE:
215 <221> NAME/KEY: CDS
216 <222> LOCATION: (1)..(1998)
218 <400> SEQUENCE: 3
219 atg agg gac cgg gtc cgc tgg cgg gtg ctt tcc ctc cct ccc ctc gcc      48
220 Met Arg Asp Arg Val Arg Trp Arg Val Leu Ser Leu Pro Pro Leu Ala
221 1          5          10          15
223 cag tgg cgg gag gtg atg gcg gcc ttg gag gtg ggg ccg gag gcc gcc      96

```

## RAW SEQUENCE LISTING

DATE: 09/28/2004

PATENT APPLICATION: US/09/938,901A

TIME: 10:38:43

Input Set : A:\11283-013001.ST25.txt

Output Set: N:\CRF4\09282004\I938901A.raw

224	Gln	Trp	Arg	Glu	Val	Met	Ala	Ala	Leu	Glu	Val	Gly	Pro	Glu	Ala	Ala	
225				20					25				30				
227	ctg	gcc	tac	tgg	cac	cgg	ggc	ttt	agg	cgc	aag	gag	gac	ctg	gac	ccc	144
228	Leu	Ala	Tyr	Trp	His	Arg	Gly	Phe	Arg	Arg	Lys	Glu	Asp	Leu	Asp	Pro	
229			35					40					45				
231	ccc	ctc	gcc	ctc	ctt	ccc	ctc	aag	ggc	ctg	agg	gag	gcg	gcg	gcc	ctc	192
232	Pro	Leu	Ala	Leu	Leu	Pro	Leu	Lys	Gly	Leu	Arg	Glu	Ala	Ala	Ala	Leu	
233			50					55					60				
235	ctg	gag	gag	gcg	ctc	cgc	cag	ggg	aag	cgg	atc	cgc	gtc	cac	ggg	gac	240
236	Leu	Glu	Glu	Ala	Leu	Arg	Gln	Gly	Lys	Arg	Ile	Arg	Val	His	Gly	Asp	
237	65					70				75					80		
239	tac	gac	gcc	gac	ggg	ctc	acg	ggc	acg	gcc	atc	ctg	gtt	cgg	ggc	ctc	288
240	Tyr	Asp	Ala	Asp	Gly	Leu	Thr	Gly	Thr	Ala	Ile	Leu	Val	Arg	Gly	Leu	
241				85					90					95			
243	gcc	gcc	ttg	ggc	gcc	gac	gtc	cac	ccc	ttc	atc	ccc	cac	cgg	ctg	gag	336
244	Ala	Ala	Leu	Gly	Ala	Asp	Val	His	Pro	Phe	Ile	Pro	His	Arg	Leu	Glu	
245			100						105					110			
247	gaa	ggg	tac	ggg	gtg	ctg	atg	gag	cgg	gtt	ccc	gag	cac	ctc	gag	gcc	384
248	Glu	Gly	Tyr	Gly	Val	Leu	Met	Glu	Arg	Val	Pro	Glu	His	Leu	Glu	Ala	
249			115					120					125				
251	tcg	gac	ctc	ttc	ctc	acc	gtg	gac	tgc	ggg	atc	acg	aac	cac	gcc	gag	432
252	Ser	Asp	Leu	Phe	Leu	Thr	Val	Asp	Cys	Gly	Ile	Thr	Asn	His	Ala	Glu	
253			130				135						140				
255	ctc	agg	gag	ctt	ttg	gaa	aac	ggg	gtg	gag	gtg	atc	gtc	acc	gac	cac	480
256	Leu	Arg	Glu	Leu	Leu	Glu	Asn	Gly	Val	Glu	Val	Ile	Val	Thr	Asp	His	
257	145					150				155					160		
259	cac	acc	ccc	ggc	aag	acc	cct	tcc	ccc	ggc	ctc	gtg	gtc	cac	ccc	gcc	528
260	His	Thr	Pro	Gly	Lys	Thr	Pro	Ser	Pro	Gly	Leu	Val	Val	His	Pro	Ala	
261				165					170					175			
263	ctc	acc	ccg	gac	ctt	aag	gag	aag	ccc	acg	ggg	gcg	ggg	gtg	gtc	ttc	576
264	Leu	Thr	Pro	Asp	Leu	Lys	Glu	Lys	Pro	Thr	Gly	Ala	Gly	Val	Val	Phe	
265			180						185					190			
267	ctc	ctc	ctc	tgg	gcc	ctc	cac	gag	cgc	ctg	ggc	ctt	ccc	cca	ccc	ctg	624
268	Leu	Leu	Leu	Trp	Ala	Leu	His	Glu	Arg	Leu	Gly	Leu	Pro	Pro	Pro	Leu	
269			195					200					205				
271	gag	tac	gcc	gac	ctc	gcc	gcg	gtg	ggc	acc	atc	gcc	gac	gtg	gcc	ccc	672
272	Glu	Tyr	Ala	Asp	Leu	Ala	Ala	Val	Gly	Thr	Ile	Ala	Asp	Val	Ala	Pro	
273			210				215						220				
275	ctt	tgg	ggc	tgg	aac	cgg	gcc	ttg	gtg	aag	gag	ggc	ctg	gcc	cgc	atc	720
276	Leu	Trp	Gly	Trp	Asn	Arg	Ala	Leu	Val	Lys	Glu	Gly	Leu	Ala	Arg	Ile	
277	225					230				235					240		
279	ccc	gcc	tcc	tcc	tgg	gtt	ggg	ctc	agg	ctt	ctg	gcc	gag	gcg	gtg	ggg	768
280	Pro	Ala	Ser	Ser	Trp	Val	Gly	Leu	Arg	Leu	Leu	Ala	Glu	Ala	Val	Gly	
281				245					250					255			
283	tac	acg	ggg	aag	gcg	gtg	gag	gtg	gcc	ttc	cgc	atc	gcc	ccc	cgg	atc	816
284	Tyr	Thr	Gly	Lys	Ala	Val	Glu	Val	Ala	Phe	Arg	Ile	Ala	Pro	Arg	Ile	
285			260					265					270				
287	aac	gcg	gca	agc	cgc	ctc	ggg	gag	gct	gag	aag	gcc	cta	agg	ctc	ctc	864
288	Asn	Ala	Ala	Ser	Arg	Leu	Gly	Glu	Ala	Glu	Lys	Ala	Leu	Arg	Leu	Leu	

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/938,901A

DATE: 09/28/2004

TIME: 10:38:43

Input Set : A:\11283-013001.ST25.txt

Output Set: N:\CRF4\09282004\I938901A.raw

289	275	280	285	
291 ctc acc gac gac gcg gcc gag gcc cag gcc ctc gtg ggg gaa ctc cac				912
292 Leu Thr Asp Asp Ala Ala Glu Ala Gln Ala Leu Val Gly Glu Leu His				
293 290	295	300		
295 cgg ctg aac gcc cgc cgc cag acc ctg gag gag gcc atg ctc agg aag				960
296 Arg Leu Asn Ala Arg Arg Gln Thr Leu Glu Glu Ala Met Leu Arg Lys				
297 305	310	315	320	
299 ctc ctc ccc cag gcg gac ccc gag gcc aag gcc atc gtc ctc ctg gac				1008
300 Leu Leu Pro Gln Ala Asp Pro Glu Ala Lys Ala Ile Val Leu Leu Asp				
301 325	330	335		
303 ccc gag ggg cac ccg ggg gtg atg ggc atc gtg gcg agc cgc atc ctg				1056
304 Pro Glu Gly His Pro Gly Val Met Gly Ile Val Ala Ser Arg Ile Leu				
305 340	345	350		
307 gag gcc acc ctc cgg ccc gtc ttc ctg gtg gcc cag ggc aag ggg acg				1104
308 Glu Ala Thr Leu Arg Pro Val Phe Leu Val Ala Gln Gly Lys Gly Thr				
309 355	360	365		
311 gtg cgg agc ctc gcc ccc atc agc gcc gtg gag gcc cta agg agc gcc				1152
312 Val Arg Ser Leu Ala Pro Ile Ser Ala Val Glu Ala Leu Arg Ser Ala				
313 370	375	380		
315 gag gac ctt ttg ttg cgc tac ggg ggg cac aag gag gcg gcg ggc ttc				1200
316 Glu Asp Leu Leu Leu Arg Tyr Gly Gly His Lys Glu Ala Ala Gly Phe				
317 385	390	395	400	
319 gcc atg gac gag gcc ctc ttc ccc gcc ttc aag gcc cgg gtg gag gcc				1248
320 Ala Met Asp Glu Ala Leu Phe Pro Ala Phe Lys Ala Arg Val Glu Ala				
321 405	410	415		
323 tac gcc gcc cgc ttc ccc gac ccc gtg cgc gag gtg gcc ctt ttg gac				1296
324 Tyr Ala Ala Arg Phe Pro Asp Pro Val Arg Glu Val Ala Leu Leu Asp				
325 420	425	430		
327 ctg ctt ccg gag ccc ggc ctc ctc ccc cag gtc ttc cgg gag ctc gcc				1344
328 Leu Leu Pro Glu Pro Gly Leu Leu Pro Gln Val Phe Arg Glu Leu Ala				
329 435	440	445		
331 ctt ttg gag ccc tac ggc gag gga aac ccc gag ccc ctc ttc ctc ctc				1392
332 Leu Leu Glu Pro Tyr Gly Glu Gly Asn Pro Glu Pro Leu Phe Leu Leu				
333 450	455	460		
335 ttc ggc gcc ccg gag gag gcc cgg cgc ctc ggg gag ggc cgc cac ctc				1440
336 Phe Gly Ala Pro Glu Glu Ala Arg Arg Leu Gly Glu Gly Arg His Leu				
337 465	470	475	480	
339 gcc ttc cgc ctg aag ggg gtg cgg gtc ctg gcc tgg aaa cag ggg gac				1488
340 Ala Phe Arg Leu Lys Gly Val Arg Val Leu Ala Trp Lys Gln Gly Asp				
341 485	490	495		
343 ctc gcc ctg ccc ccg gag gtg gag gtg gcg ggc ctc ctc agc gaa aac				1536
344 Leu Ala Leu Pro Glu Val Glu Val Ala Gly Leu Leu Ser Glu Asn				
345 500	505	510		
347 gcc tgg aac ggc cac ctc gcc tac gag gtc cag gcg gtg gac ctg cga				1584
348 Ala Trp Asn Gly His Leu Ala Tyr Glu Val Gln Ala Val Asp Leu Arg				
349 515	520	525		
351 aag cca gag gcg ctg gag ggc ggg atc gcg ccc ttc gcc tac ccc ctg				1632
352 Lys Pro Glu Ala Leu Glu Gly Gly Ile Ala Pro Phe Ala Tyr Pro Leu				
353 530	535	540		

RAW SEQUENCE LISTING ERROR SUMMARY      DATE: 09/28/2004  
PATENT APPLICATION:    US/09/938,901A      TIME: 10:38:44

Input Set : A:\11283-013001.ST25.txt  
Output Set: N:\CRF4\09282004\I938901A.raw

**Please Note:**

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:19; N Pos. 26

**VERIFICATION SUMMARY**

PATENT APPLICATION: US/09/938,901A

DATE: 09/28/2004

TIME: 10:38:44

Input Set : A:\11283-013001.ST25.txt

Output Set: N:\CRF4\09282004\I938901A.raw

L:1408 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:19 after pos.:0